
Praktikum: Business Analytics, Machine Learning and Artificial Intelligence

Umer Khan

**Information Systems and Machine Learning Lab (ISMLL)
University of Hildesheim, Germany**

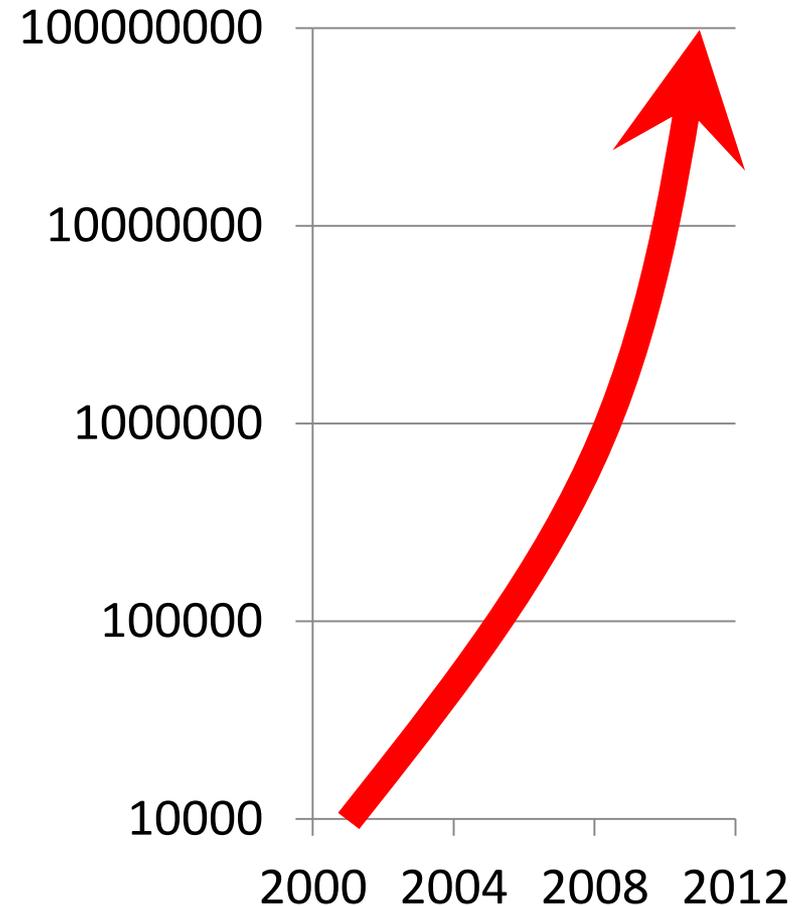
- Parallel Machine Learning Systems
 - Frameworks and algorithms

- Predicting Flight Delays

- Predict Loan Defaulters

For more topics, or topics of your interest (in Machine Learning), please talk to me personally.

- Why we need ‘*Big Data Analysis*’?
 - A popular research field these days !!
- Data hypergrowth: an example
 - Reuters-21578: about 10K docs (ModApte)
 - RCV1: about 807K docs
 - LinkedIn job title data: about 100M docs



➤ New age of big data

The world has gone mobile

5 billion cellphones produce daily data

Social networks have gone online

Twitter produces 200M tweets a day

Crowdsourcing is the reality

Labeling of 100,000+ data instances is doable

Within a week 😊

- How to manage immense amounts of data quickly, more precisely and more intelligently?

- **Praktikum Objectives:**
 - Gain hands on experience with one of the following state-of-the-art parallel machine learning systems:
 - MapReduce
 - Hadoop
 - GraphLab

 - Gain detailed knowledge of the system, by developing a parallel version of a standard machine learning algorithm like K-Means.

- Airlines are eagerly looking for intelligent solutions to optimize the flight delays
- Learning an efficient prediction model from flight data can provide important insights into causes of delays.
- **Praktikum Objectives:**
 - Using data from a running competition on *kaggle.com*, develop a machine learning algorithm to predict flight delays and identify factors causing these delays.
 - Predict whether an individual flight will be delayed?

- A very popular and demanding application area for machine learning algorithms is *Risk Analysis* for banking investments.
- With historical data of customers and their pay-back history, banks want to predict credit rating of a customers very *accurately*
- **Praktikum Objectives:**
 - Provided a data set from a *kaggle* competition, build a model that can predict the probability that a customer can fail to pay back in next two years.

- **Today:** Choose topics, and make groups
- **18.11.2013:** First idea talk
- **Weekly:** Please discuss your progress/problems. **(MUST)**
- **Final Talk:** Date will be announced shortly. (*Will be around term end*)
- **Final Report:** Approx. 20 page report.

(Important: your final grade includes the evaluation of this report)

- **Groups:** 1 – 2 students per topic.
- **Misc:**

Thanks

My Contact: khan@ismll.de

Room: C40